DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

Deep Learning Based on Attention in Semantic Segmentation: An Introductory Survey

Muna Khalaf[1], Ban N. Dhannoon[2]
[1]Computer Sciences Department, College of Science for Women, University of Bagahdad, Bagahdad, Iraq
[2]Computer Sciences Department, College of Science, Al- Nahrain University, Baghdad, Iraq
[1]munakd_comp@csw.uobaghdad.edu.iq, [2]ban.n.dhannoon@nahrainuniv.edu.iq

Abstract— *Semantic segmentation refers to labeling each pixel in the scene to its belonging object. It is a critical task for many computer vision applications that requires scene understanding because It attempts to mimic human perceptual grouping. Despite the unremitting efforts in this field, it is still a challenge and preoccupies of researchers. Semantic segmentation performance improved using deep learning rather than traditional methods. Semantic segmentation based on deep learning models requires capturing local and global context information, where deep learning models usually can extract one of them but is challenging to integrate between them. Deep learning based on attention mechanisms can gather between the capturing of local and glopal information, so it is increasingly employed in semantic segmentation. This paper gives an introductory survey of the rising topic attention mechanisms in semantic segmentation. At first, it will discuss the concept of attention and its integration with semantic segmentation requirements. Then, it will review deep learning based on attention mechanisms in semantic segmentation.*

Index Terms— *attention concept, computer vision, deep learning, semantic segmentation.*

I. INTRODUCTION

Semantic segmentation, scene labeling, or pixel-wise prediction refer to the same task: It assigns a semantic label to each pixel of an image see *Fig. 1*. It is a fundamental and challenging problem in computer vision. It is fundamental because it benefits many applications in computer vision like self-driving vehicles[1],[2] , pedestrian detection[3], [4], defect detection[5],[6], and medical diagnosis[7],[8]. Semantic segmentation is challenging because it requires both semantic and spatial accuracy[9],[10].

Semantic segmentation differs from classification because it requires classifying each pixel in the image, not only the base class of an image[11]. It differs from object localization[12] because it entails localizing all objects in an image, not only primary objects. It differs from object detection [13] because it detects all borders of objects in the image, not only detecting and bounds the primary object. Because of these differences and requirements, semantic segmentation is a complex and critical task.

Before the revolutionary deep learning era, traditional methods were employed in semantic segmentation [14], [15]. The success of deep learning in solving different computer vision problems[16]-[22] encouraged using it with semantic segmentation. Deep learning was used in semantic segmentation were led to a boom in its performance [23]. Convolutional neural networks have improved semantic segmentation performance because pre-trained features give better results than hand-crafted features [24]-[26].

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

Semantic segmentation based on deep learning faces Fundamental difficulties: it needs an unmanageably extensive convolutional network, consecutive pooling operations or convolution striding reduces feature resolution and localization accuracy, multiple scales of object existence, and it requires global and local context information [24], [25],[26], [27], [28].

FIG. 1. AN EXAMPLE OF A SEMANTIC SEGMENTATION TASK [29].

Many types of deep learning architectures are proposed to deal with previously mentioned difficulties. Fully convolutional networks (FCN) [21] induced the growth of deep convolution neural networks(DCNN) based methods that are utilized for semantic segmentation and have shown surprising performance improvements [30], [31].

FCN adapted classification networks into the segmentation task adding skip connection to enhance feature extracted, which concatenate shallow and high-level layers that integrate spatial and semantic features to improve semantic segmentation performance [23]. U-Net [32] and its later extensions [33]-[35] mainly depend on u-shaped architecture where it consists of a down-sampling (encoder)path and an up-sampling (decoder) path that uses skip connection from the corresponding down-sampling feature map. This architecture allows the U-Net to extract more semantic information, which costs more memory for skip connection. SegNet [36] costs less memory than U-Net, although it uses encoder_ decoder architecture without the skip connection.

Many architectures based on dilated convolutions are proposed for semantic segmentation. Dilated convolution supports the exponential enlargement of the receptive field with no increase in the number of parameters or the cost of computation [1], [24], [37]. Image pyramid [27],[38],[39], and Feature Pyramid [40]-[42] are employed in semantic segmentation to aggregate features at different scales. Wherein image pyramid methods feed the multi-scaled versions of the input image to the network, and Feature Pyramid takes a single-scale image as input and outputs multiple levels feature maps in a fully convolutional fashion. Deep learning based on attention mechanisms achieved successes widely in computer vision[40],Many researching areas are employed attention mechanisms like: machine translation[41], natural language processing(NLP)tasks [43], object detection[44], classification[45], robotics[46], Super-resolution imaging [47]. Semantic segmentation is one of the research areas that lately benefited from deep learning based on attention mechanisms. It appears a promising area in semantic segmentation where it improved semantic segmentation performance[9],[48]. Attention mechanisms achieve what semantic segmentation needs. It aggregates local and global contextual information[9], [48].

Received 15/February/2022; Accepted 01/May/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

This introductory survey discusses and lists deep learning based on attention mechanisms used in semantic segmentation. At first, it will introduce the attention concept and its roots in computer vision and then focus on how it is employed in semantic segmentation. Some surveys presented attention mechanisms in different research areas, like in[43], which discusses the various attention mechanisms employed to handle NLP tasks. The survey proposed in[49] provides an introductory summary of attention mechanisms in computer vision. As far as we know, no survey expanded discussed this area in semantic segmentation, Although there are brief references to it in some other surveys[50],[51].

II. ATTENTION MECHANISMS

Attention is the main characteristic of all perceptual and cognitive operations. All senses of humans employ the attention concept: for example, in a noisy room, a person can listen and understand whose talking with him. In a crowded place, a person can get attention selectively to what he needs at that moment and ignore all other details [52].

Due to limited resources, attentional mechanisms select and focus on the information most relevant for behavior [53]. At every moment, a human is surrounded by hundreds of information that cannot be processed simultaneously. So the attention concept is helped humans to balance they are processing resources with coming information by choosing the most relevant information at that moment [52].

Visual attention refers to the attention mechanisms in the visual system. The primate visual system utilizes an attention mechanism, where it can think attention as fuse visual features into relatively long-stable explanations of objects [54].Psychologists [55], [56] and Neurophysiologists [57], [58] have studied visual attention. each of them has studied visual attention from his side of science. They are studying a set of cognitive operations that decide the relevant and irrelevant information from visual scenes. Guided by these studies, computer vision scientists [59] and roboticists [60] have tried to model attention.

These efforts led to a good integration of the attention with the deep learning models [61], so it led to growing interest in deep learning based on attention mechanisms that improve results and save computation cost in many research areas in computer vision [49].

III. ATTENTION IN SEMANTIC SEGMENTATION

Since the revolutionary approach of the fully convolutional network (FCN) in semantic segmentation, extensive endeavors based on deep convolutional neural networks have been presented [51]. They aim to produce a more accurate feature map that leads to a more accurate result. Attention mechanisms in deep learning are a new class of neural networks [62] that aims to aggregate the context information and employ this information with spatial information from regular convolution layers to get more accurate semantic segmentation.

“Attention Is All You Need” that title is what researchers chose for their research [63], which proved its credibility, where attention presented there became a revolutionary jump in translation and NLP later [43]. After that, the attention mechanism integrates with deep learning in many computer vision problems [44]-[47]. Semantic segmentation used this integration as compatible with its requirement and presented many variations of attention mechanisms[64]-[67]. It can divide the attention mechanisms in semantic segmentation into

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

two models: self-attention executes in one phase, and dual attention executes within two phases.

A. Self Attention

Self Attention(SA) aggregates the context of one position from its other dependence positions that constitute the input data (e.g., a sentence or an image) no matter what distance among them [46], [48], [68]. Although improvements have occurred when using Self-attention, it is based on the simple idea presented in [57]. Equation 1 represents this idea where Attention is mapping query (Q), key(K), and value(V) to output.

$$Attention(Q,K,V) = Softmax\left(\frac{QK}{\sqrt{dk}}\right)V \qquad (1)$$

What Q, K, and V represent depends on application applies the self-attention, dk is the dimension of K. Non-local block [68] is a merging Equation 1 with non-local means algorithm [62], where it is first applying the self-attention in computer vision. The non-local block is a flexible structure that can easily combine with any existing deep learning architecture see *Fig. 2*. These simple blocks improved the baseline, but they were costly in semantic segmentation tasks because they usually require high-resolution input [69]. Self-attention is essentially an affinity matrix generated by calculating the interdependence between each pixel and other pixels; it is costly in higher resolution input. In general, the improvements in the primary form of attention progressed in two directions: capturing richer contextual information and decreasing computational complexity.

Many researchers have proposed various methods To get richer context information from self-attention, where context information is a primary key to good semantic segmentation performance. Several authors [70]-[74] used self-attention in more than one place in the deep network to boost context information from different layers or scales. In [66], it was shown that self-attention allows capturing contextual information from co-occurrent features by applying it over co-occurrent probabilities. In [57], it was shown that using self-attention over soft object regions, which is the regions of the same category, instead of over the whole scene enhances capturing context information, reducing noisy and redundant features.

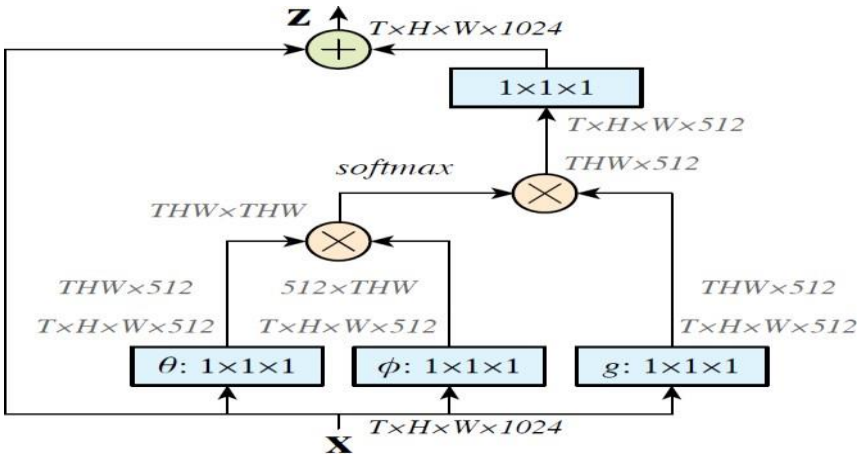

FIG. 2. NON-LOCAL BLOCK, T×H×W IS FEATURE MAPS FOR CHANNELS (1024 OR 512), "⊗" AND "⊕" SYMBOLIZE MATRIX MULTIPLICATION AND ELEMENT-WISE SUM. 1×1×1 SYMBOLIZE 1×1×1 CONVOLUTIONS [68].

Received 15/February/2022; Accepted 01/May/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

The drawback of self-attention in semantic segmentation is computation cost and memory consumption. Self-attention requires a dense relation matrix that costs computation and memory in high-resolution inputs usually required in semantic segmentation problems. To decrease the computation cost of self-attention Interlaced sparse self-attention was used [65],[75]. It used sparse relation rather than the dense relation in an array of self-attention, see *Fig. 3*. Sparse relation follows the interlacing method to evaluate the relation among pixels. [67] Interlaced Sparse with region-wise attention that approved capturing long-range contextual category information.

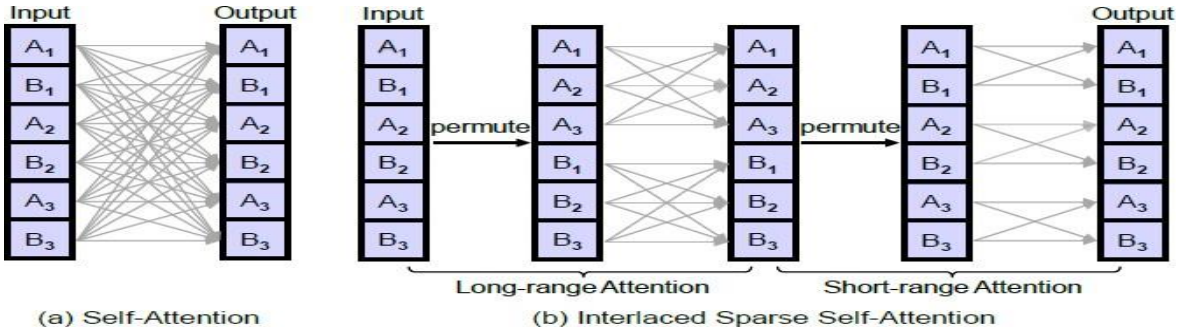

FIG. 3. INFORMATION PROPAGATION PATH OF (A) SELF-ATTENTION AND (B) INTERLACED SPARSE SELF-ATTENTION [69].

B. Dual Attention

Dual attention(DA) was presented to improve capturing long-range contextual information prepared to enhance semantic segmentation results. Dual attention captures contextual information within two parallel operations: position attention and channel attention [9] ,[76], see *Fig. 4*. position attention captures the long-range spatial contexts among positions of the features map. while channel attention captures global long-range contexts among channels [9] , [77],[78]. Although the dual attention idea of splitting the capturing of context information with two phases improves accuracy, that is expensive at computation and memory[9] , [77], [79].

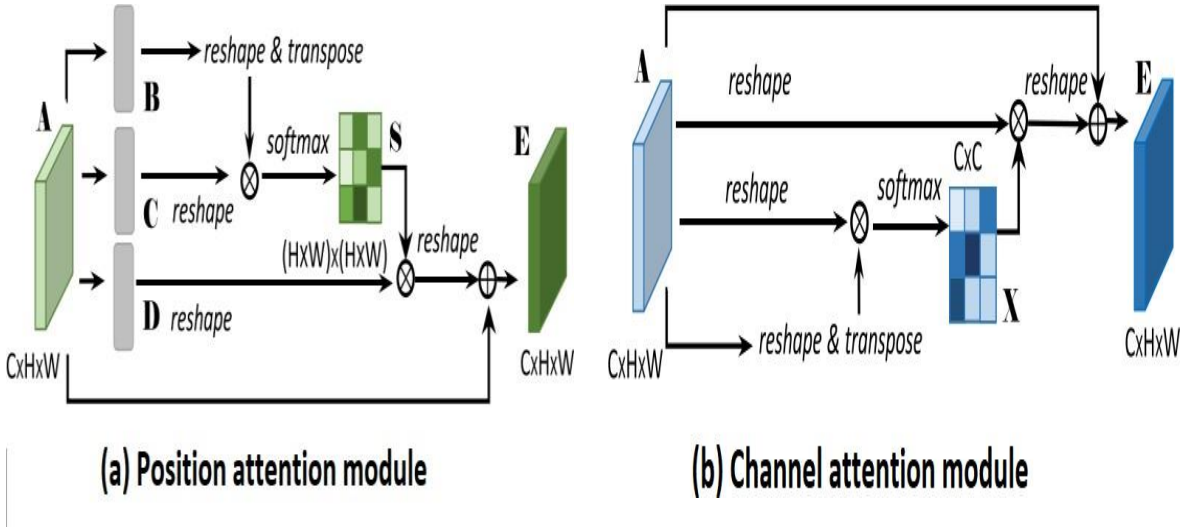

FIG. 4. THE DETAILS OF PAM AND CAM [9].

Received 15/February/2022; Accepted 01/May/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

The position attention module (PAM) [9] aggregates the spatial contexts by generating a spatial attention matrix that computes how each position impacts another position in the feature map. Some research used only position attention and got global context (channel attention goal) from the other parts of the deep network. These decrease computational complexity and give more flexibility to the dual idea in capturing context information [80],[81].

The channel attention focuses on channel interdependencies [9]. It generates a channel map representing the degree of interdependencies between channels, filtering channels to get the most discriminative channel [81]. In general, it can divide the channel attention method into two different structures: first channel attention module (CAM) proposed in [9], second Squeeze-and- Excitation module(SE) proposed in [81].

Usually, the CAM integrates with the PAM as dual attention in the networks, which are used together [9],[77], [78]. CAM evaluates the interdependencies between channel maps, pays more attention to the more informative channel, and reduces redundancy to improve semantic segmentation [79],[82].

SE recalibrates channels to pay attention to channels that have more interdependencies [83]. SE block is not expensive in computation or memory. It consists of two phases: squeeze phase uses global average pooling operation, excitation phase uses two fully connected layers followed by scaling operation [84], see *Fig. 3*. This advantage makes it a good addition for many network models. It was used with U-Net [85]-[87] , used with FCN [88],[89], Used with ResNet [84]and Used with DeepLab [90].SE improved semantic segmentation with a minimal increase in model complexity. Table I is summarized the most prominent deep network models based on attention that boosts semantic segmentation performance.

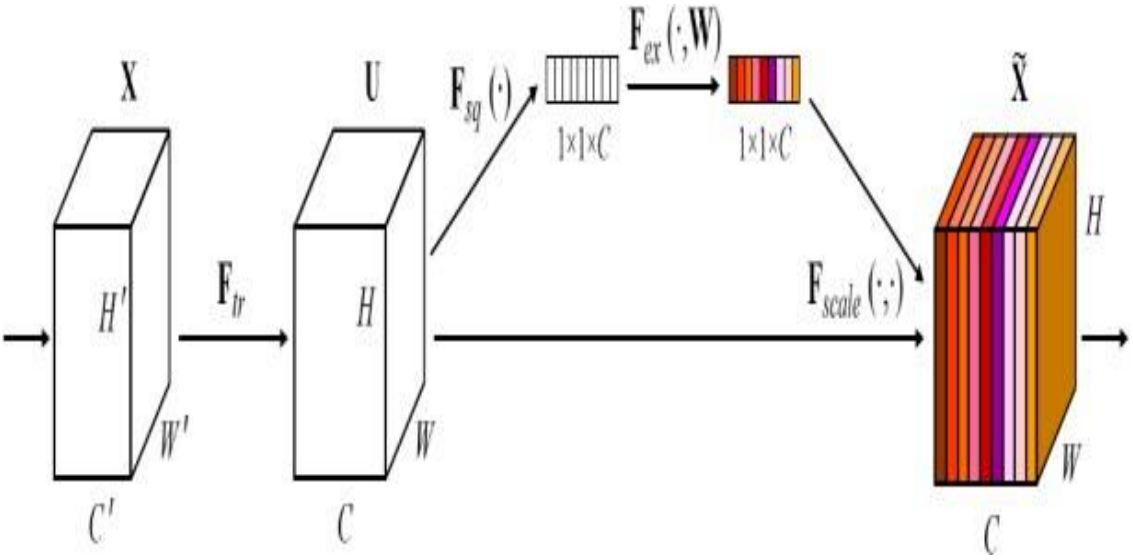

FIG. 5. A SQUEEZE-AND-EXCITATION MODULE [83].

Received 15/February/2022; Accepted 01/May/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

TABLE I. SUMMARY OF NOTABLE DEEP NETWORK MODELS EMPLOYED ATTENTION MECHANISMS IN SEMANTIC SEGMENTATION

Name of model	Year of publication	Type of attention	Contribution on Attention
OCNet[65]	2018	SA	Interlaced sparse self- attention(reduced complexity)
DAF Model [74]	2018	SA	Deep attentional Features(suppressing noise at shallow layers, add more details into features at deep layers.)
DANet[9]	2019	DA	Dual attention network(capturecontext information within two-phase, increase accuracy)
CFNet [66]	2019	SA	Build Aggregated Co-occurrent Feature(capture the context-aware information through the co- occurrent features)
USE-Net [85]	2019	SE	Improving UNET model by using SE
CANet [78]	2020	DA	Co-attention Network (color and Depth features fused in dual attention net)
NSE-Deeplab[90]	2020	SE	Improving Deeplab model by using SE
RSANet[48]	2021	SA	Regional Self-Attention(reduces noisy and redundant features)
HMANet [67]	2021	DA	Region-wise representations (capture long-range contextual category information, reduce computation time)
PTANet[91]	2021	DA	Triple Attention Block (using dual attention with region attention)
LAANet[92]	2022	DA	Efficient Asymmetric Bottleneck (proposed lightweight attention-guided model)
AGLNet[93]	2022	SA	global attention pooling (identifies a semantic descriptor's implicit information)
ESSNet[94]	2022	DA	self-attention distillation scheme(adaptively moves long-range context knowledge from teacher to student networks.)

IV. CONCLUSIONS

This paper is the first survey in the literature that focuses on deep learning based on attention mechanisms in semantic segmentation. The attention mechanisms improve the semantic segmentation results by capturing the local and global context information. A concise overview of the attention mechanisms is discussed that classifies mechanisms into two types: Self Attention and Dual attention.

Self Attention aggregates the context of one location from its other dependent positions that comprise the picture, regardless of the distance between them. It helps get richer context information, which improves semantic segmentation results, but it has a computational cost. Dual attention captures long-contextual information by splitting attention into positions and channeling attention. It improves semantic segmentation results but also costs computation and memory.

Attention mechanisms are a good and simple addition to many semantic segmentation based on deep learning models that do not require a change in the basic model. It is integrated with semantic segmentation requirements and boosts performance, dividing basic variations of attention mechanisms employed in semantic segmentation. A summary of notable deep network models that employed attention mechanisms to boost semantic segmentation is shown In Table I.

Received 15/February/2022; Accepted 01/May/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

REFERENCES

[1] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv Prepr. arXiv*, vol. 40, no. 4, pp. 834–848, 2016, doi: 10.1109/TPAMI.2017.2699184.

[2] D. Feng *et al.*, "Deep Multi-modal Object Detection and Semantic Segmentation for Autonomous Driving : Datasets , Methods , and Challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341-1360., 2020, doi: 10.1109/TITS.2020.2972974.2.

[3] T. Liu and T. Stathaki, "Faster R-CNN for Robust Pedestrian Detection Using Semantic Segmentation Network," *Front. Neurorobot.*, vol. 12, no. October, p. 64, 2018, doi: 10.3389/fnbot.2018.00064.

[4] J. Fei, W. Chen, P. Heidenreich, S. Wirges, C. Stiller, and R. O. Sep, "SemanticVoxels : Sequential Fusion for 3D Pedestrian Detection using LiDAR Point Cloud and Semantic Segmentation," *arXiv Prepr. arXiv2009.12276*, 2020.

[5] S. Mertes, A. Margraf, C. Kommer, S. Geinitz, and E. Andr, "Data Augmentation for Semantic Segmentation in the Context of Carbon Fiber Defect Detection using Adversarial Learning," *Proc. ofthe 1st Int. Conf. Deep Learn. Theory Appl. (DeLTA 2020)*, no. DeLTA, pp. 59–67, 2020, doi: 10.5220/0009823500590067.

[6] I. R. U. Xiihu *et al.*, "A Semantic Segmentation Method for Buffer Layer Defect Detection in High Voltage Cable," *E3S Web Conf.*, vol. 233, pp. 4–7, 2021.

[7] N. Alalwan, A. Abozeid, A. A. Elhabshy, and A. Alzahrani, "Efficient 3D Deep Learning Model for Medical Image Semantic Segmentation," *Alexandria Eng. J.*, vol. 60, no. 1, pp. 1231–1239, 2021, doi: 10.1016/j.aej.2020.10.046.

[8] J. Paul and C. Julien, "Deep Semantic Segmentation of Natural and Medical Images: A Review," *Artif. Intell. Rev. 54*, vol. 54, no. 1, pp. 137–178, 2021.

[9] Jun Fu *et al.*, "Dual Attention Network for Scene Segmentation," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 3146–3154, 2019.

[10] C. Peng and J. Ma, "Semantic segmentation using stride spatial pyramid pooling and dual attention decoder," *Pattern Recognit.*, vol. 107, p. 107498, 2020, doi: 10.1016/j.patcog.2020.107498.

[11] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 1161, 2007, doi: 10.1080/01431160600746456.

[12] H. Harzallah *et al.*, "Combining efficient object localization and image classification," *IEEE 12th Int. Conf. Comput. Vis.*, pp. 237–244, 2009.

[13] L. Liu *et al.*, "Deep Learning for Generic Object Detection : A Survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020, doi: 10.1007/s11263-019-01247-4.

[14] B. Li, Y. Shi, Z. Qi, and Z. Chen, "A survey on semantic segmentation," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2018-Novem, no. November, pp. 1233–1240, 2019, doi: 10.1109/ICDMW.2018.00176.

[15] F. Schroff, A. Criminisi, and A. Zisserman, "Object class segmentation using random forests," *Proc. Br. Mach. Vis. Conf.*, pp. 54.1-54.10, 2008, doi: 10.5244/C.22.54.

[16] X. Wang, "Deep learning for indoor fingerprinting using channel state information," *Wirel. Commun. Netw. Conf.*, 2015, doi: 10.1109/WCNC.2015.7127718.

[17] A. Eitel, L. Spinello, and M. Riedmiller, "Multimodal Deep Learning for Robust RGB-D Object Recognition," *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 681–687, 2015.

[18] R. Fakoor and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," *Proc. Int. Conf. Mach. Learn.*, vol. 28, no. August, pp. 3937–3949, 2017.

[19] A. A. Abdulhussein and F. A. Raheem, "Hand Gesture Recognition of Static Letters American Sign Language (ASL) Using Deep Learning," *Eng. Technol. J.*, vol. 38, no. 06, pp. 926–937, 2020.

[20] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object Detection With Deep Learning : A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[21] A. M. Salih and B. N. Dhannoon, "Color Model Based Convolutional Neural Network for Image Spam Classification," *Al-Nahrain J. Sci.*, vol. 23, no. 4, pp. 44–48, 2020, doi: 10.22401/ANJS.23.4.08.

[22] H. M. Ahmed and H. H. Mahmoud, "Effect of Successive Convolution Layers to Detect Gender," *Iraqi J. Sci.*, vol. 59, no. 3C, pp. 1717–1732, 2018, doi: 10.24996/ijs.2018.59.3C.17.

[23] J. Zhuang, J. Yang, L. Gu, and N. Dvornek, "Fully Convolutional Networks for Semantic Segmentation," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 3431-3440., 2015, doi: 10.1109/ICCVW.2019.00113.

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

[24] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018, doi: 10.1109/TPAMI.2017.2699184.

[25] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," *Eur. Conf. Comput. Vis.*, pp. 519–534, 2016, doi: 10.1007/978-3-319-46487-9_32.

[26] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv*, pp. 1–13, 2018.

[27] C. Couprie, L. Najman, and Y. Lecun, "Learning Hierarchical Features for Scene Labeling," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 35, no. 8, pp. 1915–1929, 2013, doi: 10.1109/TPAMI.2012.231.

[28] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv*, 2017.

[29] M. Cordts *et al.*, "The Cityscapes Dataset for Semantic Urban Scene Understanding," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 3213–3223, 2016.

[30] Z. Wu, C. Shen, and A. van den Hengel, "High-performance Semantic Segmentation Using Very Deep Fully Convolutional Networks," *arXiv Prepr. arXiv*, p. 1604.04339, 2016, [Online]. Available: http://arxiv.org/abs/1604.04339.

[31] D. Lin, J. Dai, and K. He, "ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation," *Proc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3159–3167, 2016, [Online]. Available: http://research.microsoft.com/en-us/um/.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net : Convolutional Networks for Biomedical," *Int. Conf. Med. image Comput. Comput. Interv. Springer, Cham*, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4.

[33] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019, doi: 10.1109/ACCESS.2019.2908991.

[34] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, and Y. Iwamoto, "Unet 3+: A full-scale connected unet for medical image segmentation," *Int. Conf. Acoust. Speech Signal Process.*, pp. 1055–1059, 2020.

[35] Z. Zhou, M. M. Rahman Siddiquee, and Tajbakhsh Nima, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Trans Med Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020, doi: 10.1109/TMI.2019.2959609.UNet.

[36] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling," 2015, [Online]. Available: http://arxiv.org/abs/1505.07293.

[37] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, 2016.

[38] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 2650–2658, 2015, doi: 10.1109/ICCV.2015.304.

[39] M. O. ´c and S. Š. ´c, "Efficient semantic segmentation with pyramidal fusion," *Pattern Recognit.*, p. 107611., 2021, doi: 10.1016/j.patcog.2020.107611.

[40] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 936–944, 2017, doi: 10.1109/CVPR.2017.106.

[41] J. Martinsson and O. Mogren, "Semantic segmentation of fashion images using feature pyramid networks," *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, pp. 3133–3136, 2019, doi: 10.1109/ICCVW.2019.00382.

[42] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," *arXiv*, pp. 6399–6408, 2019.

[43] D. Hu, "An Introductory Survey on Attention Mechanisms in NLP Problems," *Proc. SAI Intell. Syst. Conf.*, pp. 432–448, 2019.

[44] W. Li, K. Liu, L. Zhang, and F. Cheng, "Object detection based on an adaptive attention mechanism," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020, doi: 10.1038/s41598-020-67529-x.

[45] R. Xu, Y. Tao, Z. Lu, and Y. Zhong, "Attention-mechanism-containing neural networks for high-resolution remote sensing image classification," *Remote Sens.*, vol. 10, no. 10, pp. 1–29, 2018, doi: 10.3390/rs10101602.

[46] J. Yan *et al.*, "Trajectory prediction for intelligent vehicles using spatial-attention mechanism," *IET Intell. Transp. Syst.*, vol. 14, no. 13, pp. 1855–1863, 2020, doi: 10.1049/iet-its.2020.0274.

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

[47] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A Fast and Lightweight Network for Single-Image Super Resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, 2021, doi: 10.1109/TCYB.2020.2970104.

[48] D. Zhao, C. Wang, Y. Gao, Z. Shi, and F. Xie, "Semantic Segmentation of Remote Sensing Image Based on Regional Self-Attention Mechanism," *IEEE Geosci. Remote Sens. Lett.*, 2021.

[49] J. Sun, J. Jiang, and Y. Liu, "An Introductory Survey on Attention Mechanisms in Computer Vision Problems," *6th Int. Conf. Big Data Inf. Anal. 2020*, pp. 295–300, 2020, doi: 10.1109/BigDIA51454.2020.00054.

[50] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019, doi: 10.1016/j.neucom.2019.02.003.

[51] S. Hao, Y. Zhou, and Y. Guo, "A Brief Survey on Semantic Segmentation with Deep Learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020, doi: 10.1016/j.neucom.2019.11.118.

[52] S. Frintrop and E. Rome, "Computational Visual Attention Systems and their Cognitive Foundations : A Survey," vol. 7, no. 1, pp. 1–46, 2010.

[53] M. M. Chun, J. D. Golomb, and N. B. Turk-browne, "A Taxonomy of External and Internal Attention," *Annu. Rev. Psychol.*, vol. 62, pp. 73-101., 2011, doi: 10.1146/annurev.psych.093008.100427.

[54] R. A. Rensink, "The Dynamic Representation of Scenes," *Vis. cogn.*, vol. 7, pp. 17–42, 2000.

[55] C. Paper and R. A. Rensink, "How Much of a Scene is Seen ? The Role of Attention in Scene Perception.," *Investig. Ophthalmol. Vis. Sci.*, no. May, p. S707, 1997.

[56] A. Michael, G. A. Alvarez, K. N. Natural-, M. A. Cohen, G. A. Alvarez, and K. Nakayama, "Natural-Scene Perception Requires Attention," *Psychol. Sci.*, vol. 9, no. 22, pp. 1165–1172, 2011, doi: 10.1177/0956797611419168.

[57] J. C. Martinez-trujillo and S. Treue, "Feature-Based Attention Increases the Selectivity of Population Responses in Primate Visual Cortex," *Curr. Biol.*, vol. 14, no. 9, pp. 744–751, 2004, doi: 10.1016/j.

[58] M. Stalter, S. Westendorff, and A. Nieder, "Feature-based attention processes in primate prefrontal cortex do not rely on feature similarity," *Cell Rep.*, vol. 36, no. 5, p. 109470, 2021, doi: 10.1016/j.celrep.2021.109470.

[59] P. R. Roelfsema, "Cortical algorithms for perceptual grouping," *Annu. Rev. Neurosci.*, vol. 29, no. May, pp. 203–227, 2006, doi: 10.1146/annurev.neuro.29.051605.112939.

[60] M. Begum and F. Karray, "Visual Attention for Robotic Cognition : A Survey," vol. 3, no. 1, pp. 92–105, 2011.

[61] A. D. S. Correia and E. L. Colombini, "Attention, please! A survey of Neural Attention Models in Deep Learning," *arXiv Prepr. arXiv2103.16775*, 2021.

[62] A. Buades *et al.*, "A non-local algorithm for image denoising," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 60–65, 2005.

[63] A. Vaswani *et al.*, "Attention is All you Need," *Adv. neural Inf. Process. Syst.*, pp. 5998–6008, 2017.

[64] P. Tang *et al.*, "Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging," *Comput. Methods Programs Biomed.*, vol. 178, pp. 289–301, 2019, doi: 10.1016/j.cmpb.2019.07.005.

[65] Y. Yuan, L. Huang, J. Guo, C. Zhang, and C. V Mar, "Ocnet: Object context network for scene parsing," *arXiv Prepr. arXi*, p. 1809.00916, 2018.

[66] H. Zhang, H. Zhang, and C. Wang, "Co-occurrent Features in Semantic Segmentation," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 548–557, 2019.

[67] R. Niu *et al.*, "Hybrid Multiple Attention Network for Semantic Segmentation in Aerial Images," *IEEE Trans. Geosci. Remote Sensing.*, vol. 9, no. 10, p. 571, 2021.

[68] X. Wang and R. Girshick, "Non-local Neural Networks," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 7794–7803, 2018.

[69] L. Huang, Y. Yuan, J. Guo, C. Zhang, X. Chen, and J. Wang, "Interlaced Sparse Self-Attention for Semantic Segmentation," *arXiv Prepr. arXiv*, p. 1907.12273, 2019.

[70] M. Yan, J. Wang, J. Li, K. Zhang, and Z. Yang, "Traffic scene semantic segmentation using self-attention mechanism and bi-directional GRU to correlate context," *Neurocomputing*, vol. 386, pp. 293–304, 2020, doi: 10.1016/j.neucom.2019.12.007.

[71] W. Zhou, M. Ieee, J. Yuan, J. Lei, and T. Luo, "TSNet : Three-stream Self-attention Network for RGB-D Indoor Semantic Segmentation," *IEEE Intell. Syst.*, 2020, doi: 10.1109/MIS.2020.2999462.

[72] S. Images, "Self-Attention in Reconstruction Bias U-Net for Semantic Segmentation of Building Rooftops in Optical Remote Sensing Images," *Remote Sens.*, vol. 13, no. 13, p. 2524, 2021.

[73] V. Marsocci and S. Scardapane, "MARE : Self-Supervised Multi-Attention REsu-Net for Semantic Segmentation in Remote Sensing," *Remote Sens.*, vol. 13, no. 16, p. 3275, 2021.

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.1.9

[74] Y. Wang, Z. Deng, X. Hu, and L. Zhu, "Deep Attentional Features for Prostate Segmentation in Ultrasound Deep Attentional Features for Prostate Segmentation in Ultrasound," *Int. Conf. Med. Image Comput. Comput. Interv.*, no. September, pp. 523–530, 2018.

[75] M. Liu, H. Yin, and Department, "Sparse Spatial Attention Network For Semantic Segmentation," *IEEE Int. Conf. Image Process.*, pp. 644–648, 2021.

[76] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE J. Biomed. Heal. informatics*, vol. 25, no. 1, pp. 121–130, 2020.

[77] J. Li, J. Xiu, Z. Yang, and C. Liu, "Dual Path Attention Net for Remote Sensing Semantic Image Segmentation," 2020.

[78] H. Zhou, L. Qi, Z. Wan, and H. Huang, "RGB-D Co-attention Network for Semantic Segmentation," *Proc. Asian Conf. Comput. Vis.*, pp. 1–18, 2020.

[79] Y. Liu, C. Xu, Z. Chen, C. Chen, H. Zhao, and X. Jin, "Deep Dual-Stream Network with Scale Context Selection Attention Module for Semantic Segmentation," *Neural Process. Lett.*, vol. 51, no. 3, pp. 2281–2299, 2020, doi: 10.1007/s11063-019-10148-z.

[80] X. Chen *et al.*, "Semantic boundary enhancement and position attention network with long-range dependency for semantic segmentation," *Appl. Soft Comput.*, vol. 109, p. 107511, 2021, doi: 10.1016/j.asoc.2021.107511.

[81] T. Chowdhury and M. Rahnemoonfar, "Self Attention Based Semantic Segmentation on a Natural Disaster Dataset," *2021 IEEE Int. Conf. Image Process.*, pp. 2798–2802, 2021.

[82] H. Li, K. Qiu, L. Chen, X. Mei, L. Hong, and C. Tao, "SCAttNet : Semantic Segmentation Network with Spatial and Channel Attention Mechanism for High-Resolution Remote Sensing Images," vol. 14, no. 8, pp. 2–6, 2019.

[83] and G. S. Hu, Jie, Li Shen, "Squeeze-and-Excitation Networks," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 7132–7141, 2018.

[84] P. Ghosal, L. Nandanwar, and S. Kanchan, "Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network," *Second Int. Conf. Adv. Comput. Commun. Paradig.*, no. May 2020, pp. 1–6, 2019, doi: 10.1109/ICACCP.2019.8882973.

[85] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating Fully Convolutional Networks with Spatial and Channel ' Squeeze & Excitation ' Blocks," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 540–549, 2018.

[86] L. Rundo, C. Han, Y. Nagano, and J. Zhang, "USE-Net : incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing*, vol. 356, pp. 31–43, 2019.

[87] A. Lee, I. Woo, D. Kang, S. Chai, H. Lee, and N. Kim, "Fully automated segmentation on brain ischemic and white matter hyperintensities lesions using semantic segmentation networks with squeeze-and-excitation blocks in MRI," *Informatics Med. Unlocked*, vol. 21, pp. 706–714, 2020, doi: 10.1016/j.imu.2020.100440.

[88] Z. Zhong *et al.*, "Squeeze-and-Attention Networks for Semantic Segmentation," *he IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 13065–13074.

[89] V. Alves and C. A. Silva, "Adaptive feature recombination and recalibration for semantic segmentation: application to brain tumor segmentation in MRI," *IEEE Trans. Med. Imaging*, vol. 38, no. 12, pp. 2914–2925.

[90] Y. Lin, D. Xu, N. Wang, and Z. Shi, "Road Extraction from Very-High-Resolution Remote Sensing Images via a Nested SE-Deeplab Model," *Remote sensing, 12(18), p.2985.*, vol. 12, no. 18, pp. 1–20, 2020.

[91] H. Cheng, J. Lu, M. Luo, W. Liu, and K. Zhang, "PTANet : Triple Attention Network for point cloud semantic segmentation," *Eng. Appl. Artif. Intell.*, vol. 102, no. March, p. 104239, 2021, doi: 10.1016/j.engappai.2021.104239.

[92] Zhang, X., Du, B., Wu, Z. et al. LAANet: lightweight attention-guided asymmetric network for real-time semantic segmentation. Neural Comput & Applic 34, 3573–3587 (2022). https:/doi.org/10.1007/s00521-022-06932-z.

[93] Li, Jiangyun, Sen Zha, Chen Chen, Meng Ding, Tianxiang Zhang, and Hong Yu. "Attention guided global enhancement and local refinement network for semantic segmentation." IEEE Transactions on Image Processing (2022), doi: 10.1109/TIP.2022.3166673.

[94] An, Shumin, Qingmin Liao, Zongqing Lu, and Jing-Hao Xue. "Efficient Semantic Segmentation via Self-Attention and Self-Distillation." IEEE Transactions on Intelligent Transportation Systems (2022), doi: 10.1109/TITS.2021.3139001.

Received 15/February/2022; Accepted 01/May/2022
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*